МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ

Национальный исследовательский университет ИТМО

ФАКУЛЬТЕТ БЕЗОПАСНОСТИ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

Управление мобильными устройствами

Лабораторная работа №2
«Обработка и тарификация трафика NetFlow»
Вариант№15

Работу выполнил:

Студент группы №3352
Невесенко В. Н. _____

Работу проверил:

Федоров И. Р . _____

Санкт-Петербург, 2020

Цель работы: изучение технологии работы протокола NetFlow, а также разработка и реализация программного модуля обработки трафика NetFlow v5 и тарификации абонента.

Описание выбранных средств реализации и обоснования выбора:

Разработанный мною программный модуль был реализован на языке Python с использованием библиотеки обработки и анализа данных Pandas. Выбор данного языка программирования обусловлен крайне низким порогом вхождения, очень широким функционалом ввиду его «популярности» и очень динамично развивающегося сообщества, которое разрабатывает множество библиотек и модулей, а также издает множество обучающих материалов для повышения уровня владения данным языком программирования.

Библиотека Pandas выбрана, прежде всего, из-за внушительного функционала, связанного с обработкой данных. Данная библиотека позволяет оптимизировать код ввиду своих возможностей, позволяя реализовать одни и те же функции вводом значительно меньшего количества команд. Также данная библиотека позволяет добиваться высокой производительности от программы ввиду особенностей своего написания.

Также для построения и удобной реализации графиков мною были выбраны пакеты из библиотеки matplotlib – pyplot и dates. Первый нужен для непосредственно построения графика зависимости, а второй нужен для удобства отображения дат.

Для удобной обработки дампа NetFlow мною был выбран пакет numpy, для реализации перевода «сырого» дампа в файл .csv (для удобства его дальнейшей обработки), мною была задействована функция os, благодаря которой я смог реализовать выполнение терминальной команды *nfdump -r nfcapd.202002251200 -o \"fmt:%ts,%sa,%da,%ibyt,%obyt\" | sed \"s/ //g\" | ghead -n -4 > data.csv*

Исходный код:

Для оптимизации и разделения функций, были созданы три отдельных программных модуля - obrab.py, tarif.py, graphs.py. Как можно догадаться из названия, первый отвечает за обработку дампа Netflow, а также в нем реализованы функции, необходимые для дальнейшей работы, tarif.py необходим для произведения вычислений по тарификации, а последний же файл выводит график зависимости объема трафика от времени. Листинг всех программных модулей представлен ниже:

Файл obrab.py

```python
import pandas as pd
import os
import numpy as np
import matplotlib.pyplot as mpl
from matplotlib.dates import DateFormatter


nfcom = "nfdump -r nfcapd.202002251200 -o \"fmt:%ts,%sa,%da,%ibyt,%obyt\" | sed \"s/ //g\" | ghead -n -4 > data.csv"
os.system(nfcom)

def sum_traf(k, Q):
    return Q * k

def graph(times, values):
    fig, ax = mpl.subplots(figsize=[15,5])
    ax.plot(times, values)
    DF = DateFormatter("%H:%M:%S")
    ax.xaxis.set_major_formatter(DF)
    mpl.title('Зависимость объема трафика от времени')
    mpl.xlabel('Время')
    mpl.ylabel('Всего байт в пакете')
    mpl.gcf().autofmt_xdate()
    mpl.show()

IP_addr = '77.74.181.52'
koef = 1.5

rdcsv = pd.read_csv('data.csv', skiprows=1, header=None)
```

Файл tarif.py

```python
import obrab as obr
import math

rc = obr.pd.read_csv('data.csv', skiprows=1, header=None)
rc.columns = ['t', 'sa', 'da', 'ibys', 'obys']
rc = rc[obr.np.logical_or(rc.sa == obr.IP_addr, rc.da == obr.IP_addr)]
rc.ibys = rc.ibys.apply(lambda row: int(row) if 'M' not in row else (int(float(row[:-1])*10**6)))
rc.t = rc.t.apply(lambda row: row[10:18])
print(f'Всего соединений с IP {obr.IP_addr}: {rc.shape[0]} шт.\n')

outcoming_traffic = rc[rc.sa == obr.IP_addr].ibys.sum() / 10**3
incoming_traffic =  rc[rc.da == obr.IP_addr].ibys.sum() / 10**3

print(f'Входящий трафик составил {incoming_traffic:0.2f} Кб.')
print(f'Исходящий трафик составил {outcoming_traffic:0.2f} Кб.')

print(f'Счет за входящий трафик: {math.floor(obr.koef * incoming_traffic):0.2f} р.')
print(f'Счет за исходящий трафик: {math.floor(obr.koef * outcoming_traffic):0.2f} р.')
```

Файл graphs.py

```python
import obrab as obr
import tarif as trf

times = obr.np.sort(trf.rc.t.unique())
values = []
for t in times:
    values.append(trf.rc.loc[trf.rc.t == t, ['ibys', 'obys']].sum().sum())

obr.graph(obr.pd.to_datetime(times), values)
```

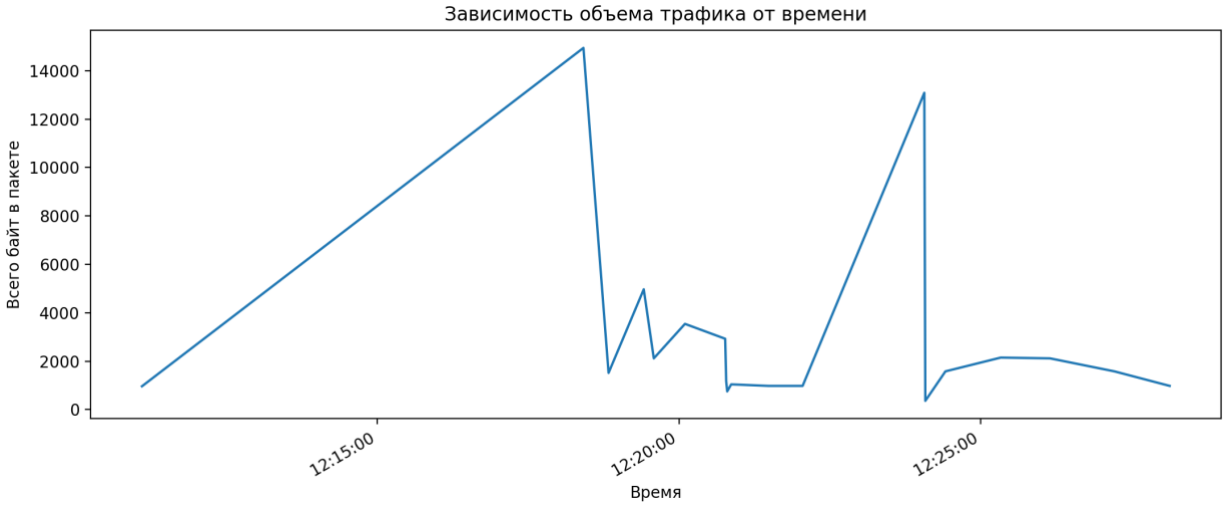
Вывод программы:

Так как объем трафика был маленький, то тарификация шла по коэффициенту 1,5 рубля за КИЛОБАЙТ. Вывод подсчета тарификации приведен ниже.

```
Всего соединений с IP 77.74.181.52: 52 шт.

Входящий трафик составил 39.45 Кб.
Исходящий трафик составил 18.28 Кб.
Счет за входящий трафик: 59.00 р.
Счет за исходящий трафик: 27.00 р.
[Finished in 1.3s]
```

Вывод зависимости представлен ниже:



Исходя из полученных данных можно проследить наиболее и наименее активное время течения трафика, следовательно, можно сделать вывод, что данный график является достаточно информативным

Выводы:

В данной Лабораторной работе были изучены технологии работы протокола NetFlow, а также разработан и реализован программный модуль обработки трафика NetFlow v5 и тарификации абонента с указанным IP-адресом.